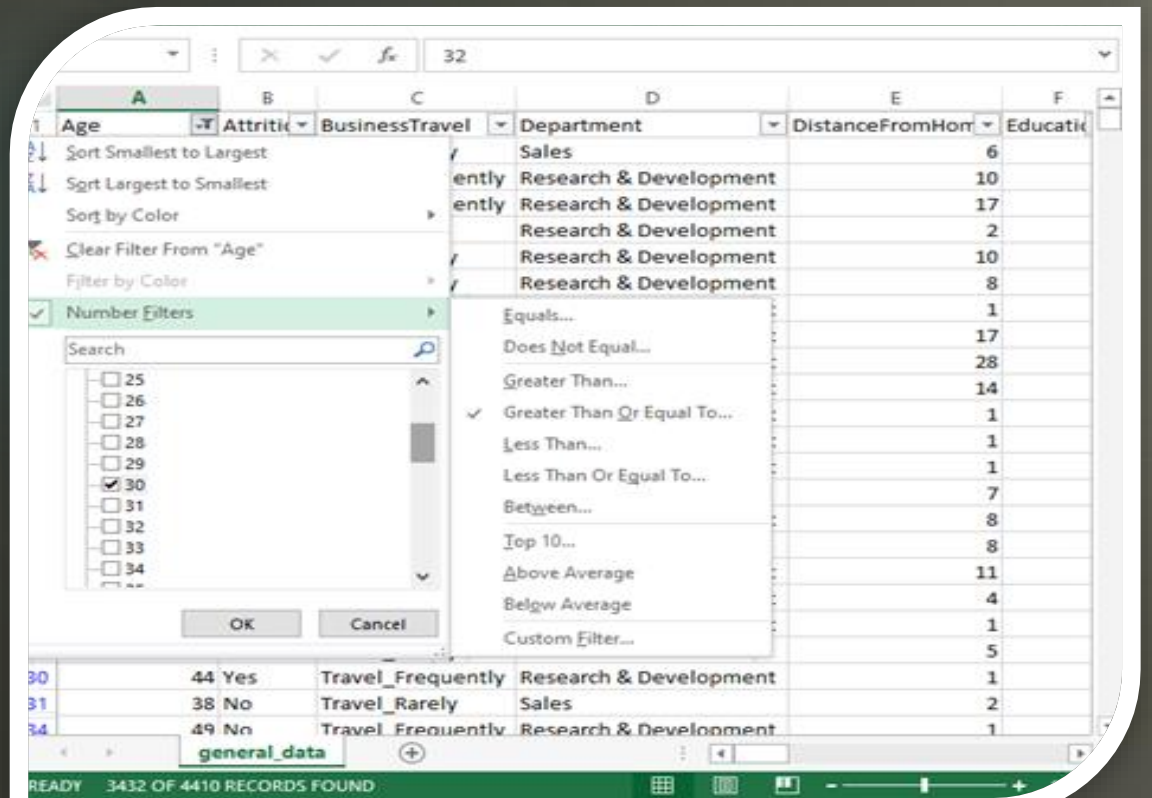
TASK 1: HR DATA ANALYSIS ASSESSMENT

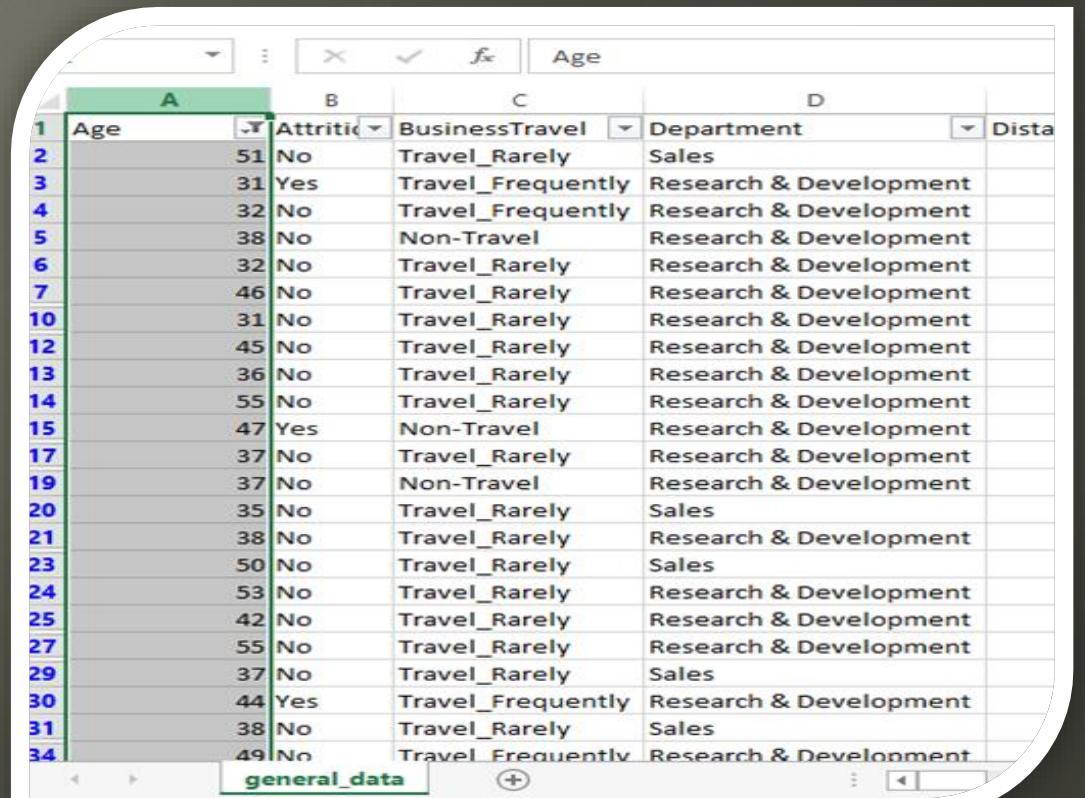
Name : Harshita Shinde

Data Analyst Intern @ PSYLIQ

1. Using Excel, how would you filter the dataset to only show employees aged 30 and above?

- Go to data tab in excel ribbon.
- Click on Filter button.A drop down arrows to the column headers will added to the dataset.
- Click on that drop-down arrow.
- Unselect the 'select all' checkbox.
- Select age 30 and then click on number filters .Select the option for 'Greater Than or equal to'.
- Now a custom auto filter will occurred . Type age 30.Click on ok.

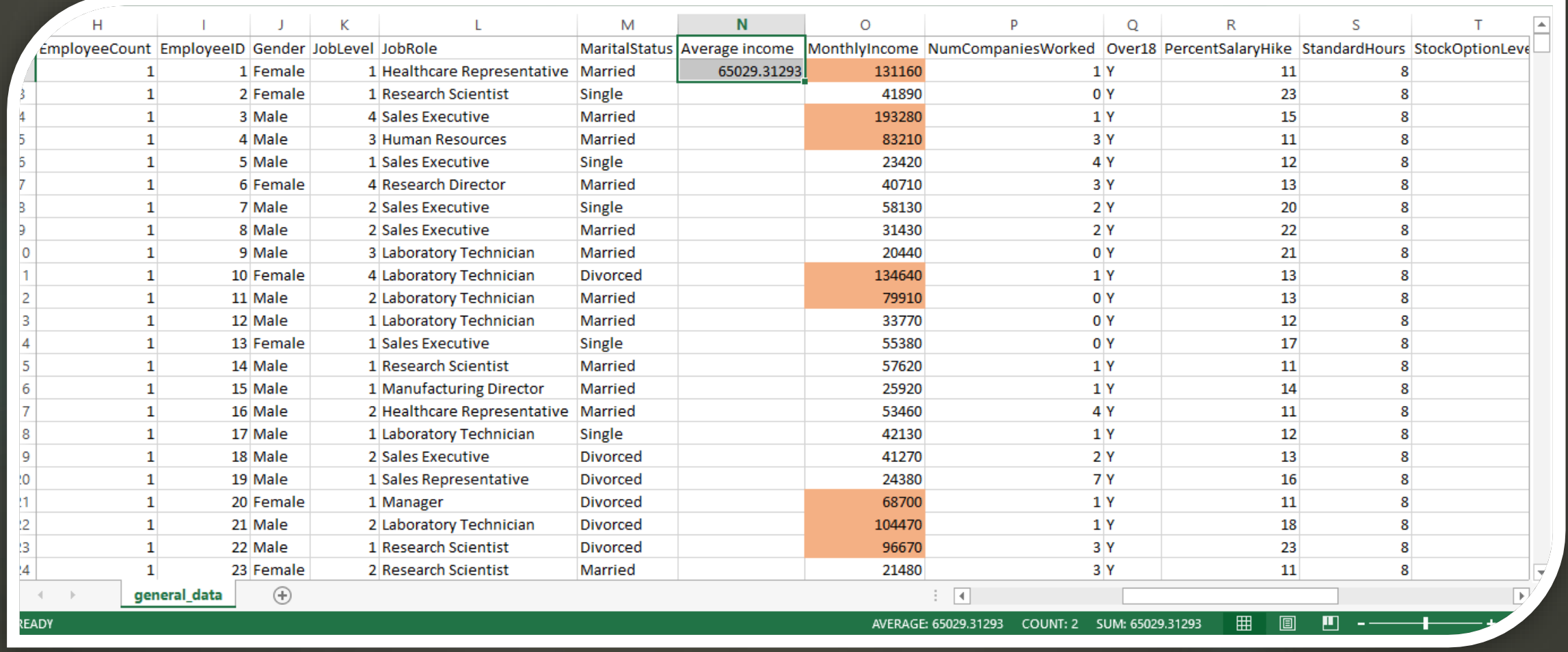




2. Create a pivot table to summarize the average Monthly Income by Job Role.

Row Labels	Average of MonthlyIncome
Healthcare Representative	60983.74046
Human Resources	58528.07692
Laboratory Technician	66314.05405
Manager	63395.88235
Manufacturing Director	69183.72414
Research Director	65473.125
Research Scientist	64975.68493
Sales Executive	65186.68712
Sales Representative	65370.96386
Grand Total	**65029.312**

3. Apply conditional formatting to highlight employees with Monthly Income above the company's average income.

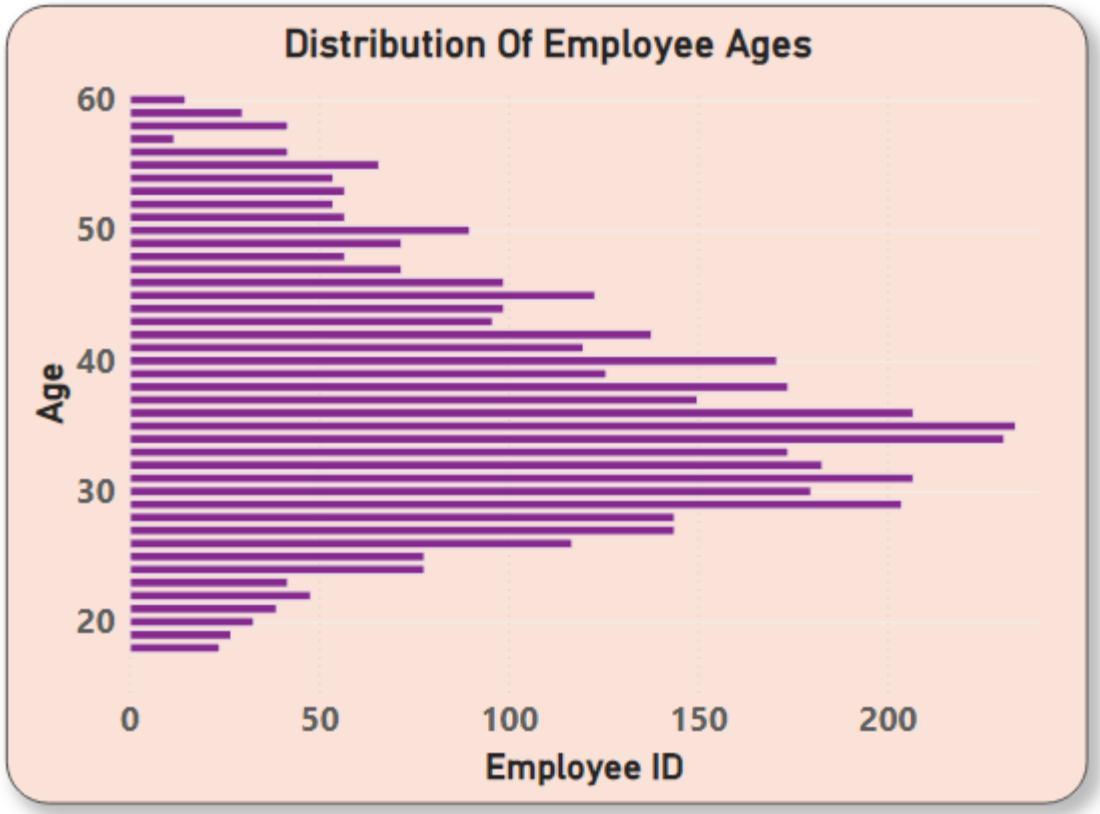


EmployeeCount	EmployeeID	Gender	JobLevel	JobRole	MaritalStatus	Average income	MonthlyIncome	NumCompaniesWorked	Over18	PercentSalaryHike	StandardHours	StockOptionLeve
1	1	Female	1	Healthcare Representative	Married	65029.31293	131160	1	Y	11	8	
1	2	Female	1	Research Scientist	Single		41890	0	Y	23	8	
1	3	Male	4	Sales Executive	Married		193280	1	Y	15	8	
1	4	Male	3	Human Resources	Married		83210	3	Y	11	8	
1	5	Male	1	Sales Executive	Single		23420	4	Y	12	8	
1	6	Female	4	Research Director	Married		40710	3	Y	13	8	
1	7	Male	2	Sales Executive	Single		58130	2	Y	20	8	
1	8	Male	2	Sales Executive	Married		31430	2	Y	22	8	
1	9	Male	3	Laboratory Technician	Married		20440	0	Y	21	8	
1	10	Female	4	Laboratory Technician	Divorced		134640	1	Y	13	8	
1	11	Male	2	Laboratory Technician	Married		79910	0	Y	13	8	
1	12	Male	1	Laboratory Technician	Married		33770	0	Y	12	8	
1	13	Female	1	Sales Executive	Single		55380	0	Y	17	8	
1	14	Male	1	Research Scientist	Married		57620	1	Y	11	8	
1	15	Male	1	Manufacturing Director	Married		25920	1	Y	14	8	
1	16	Male	2	Healthcare Representative	Married		53460	4	Y	11	8	
1	17	Male	1	Laboratory Technician	Single		42130	1	Y	12	8	
1	18	Male	2	Sales Executive	Divorced		41270	2	Y	13	8	
1	19	Male	1	Sales Representative	Divorced		24380	7	Y	16	8	
1	20	Female	1	Manager	Divorced		68700	1	Y	11	8	
1	21	Male	2	Laboratory Technician	Divorced		104470	1	Y	18	8	
1	22	Male	1	Research Scientist	Divorced		96670	3	Y	23	8	
1	23	Female	2	Research Scientist	Married		21480	3	Y	11	8	

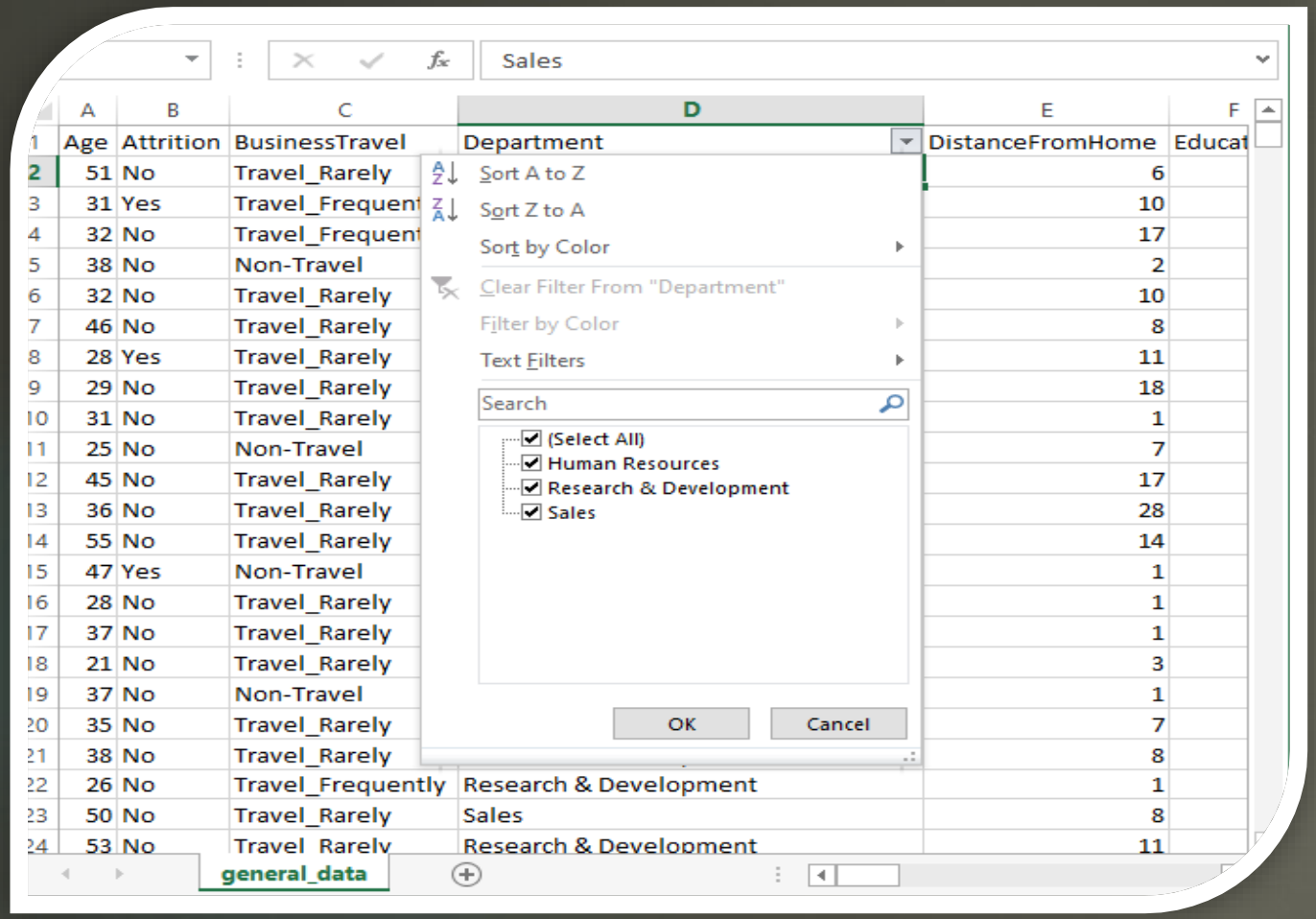
general_data

READY AVERAGE: 65029.31293 COUNT: 2 SUM: 65029.31293

4. Create a bar chart in Excel to visualize the distribution of employee ages.



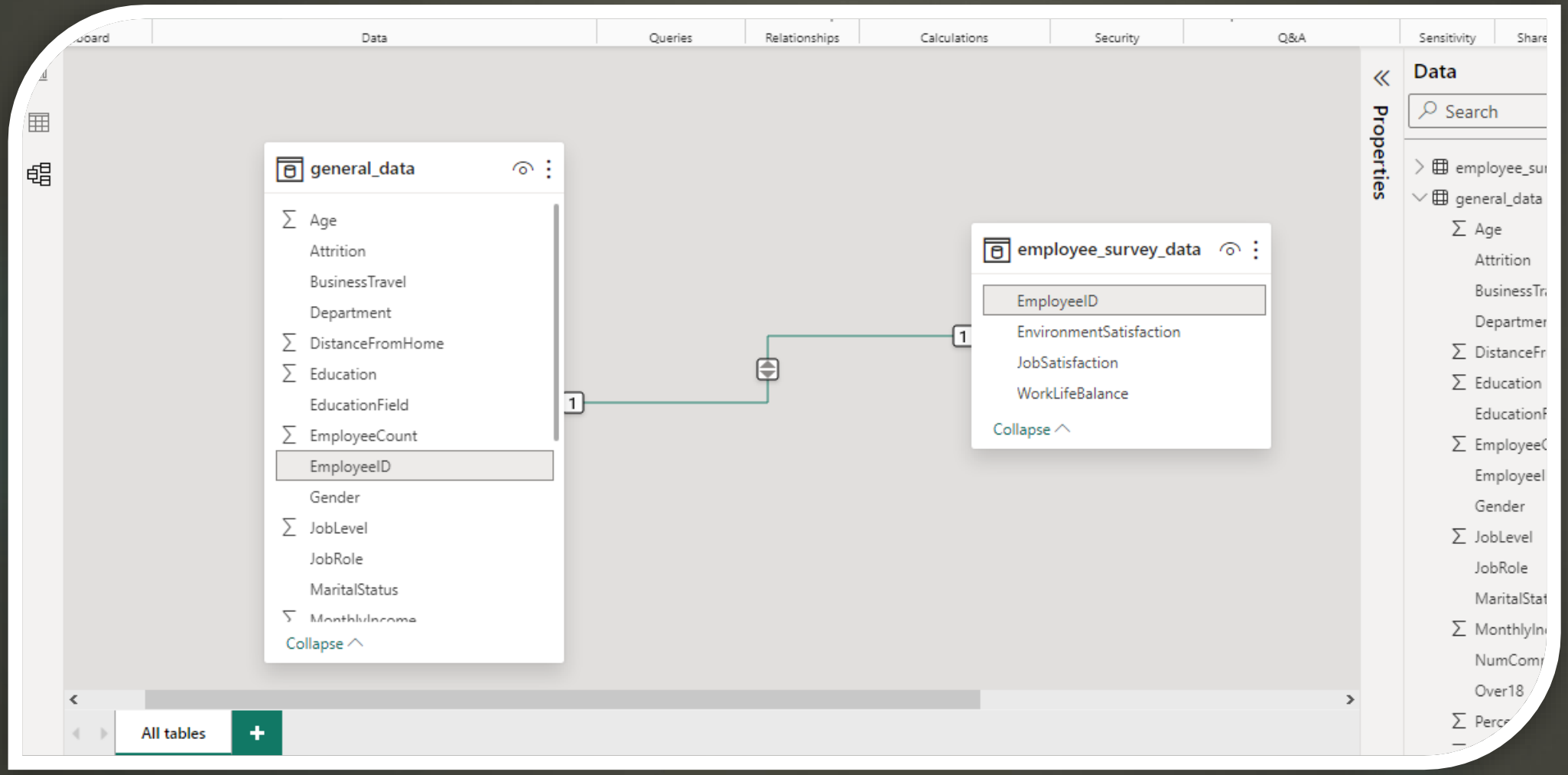
5. Identify and clean any missing or inconsistent data in the "Department" column.



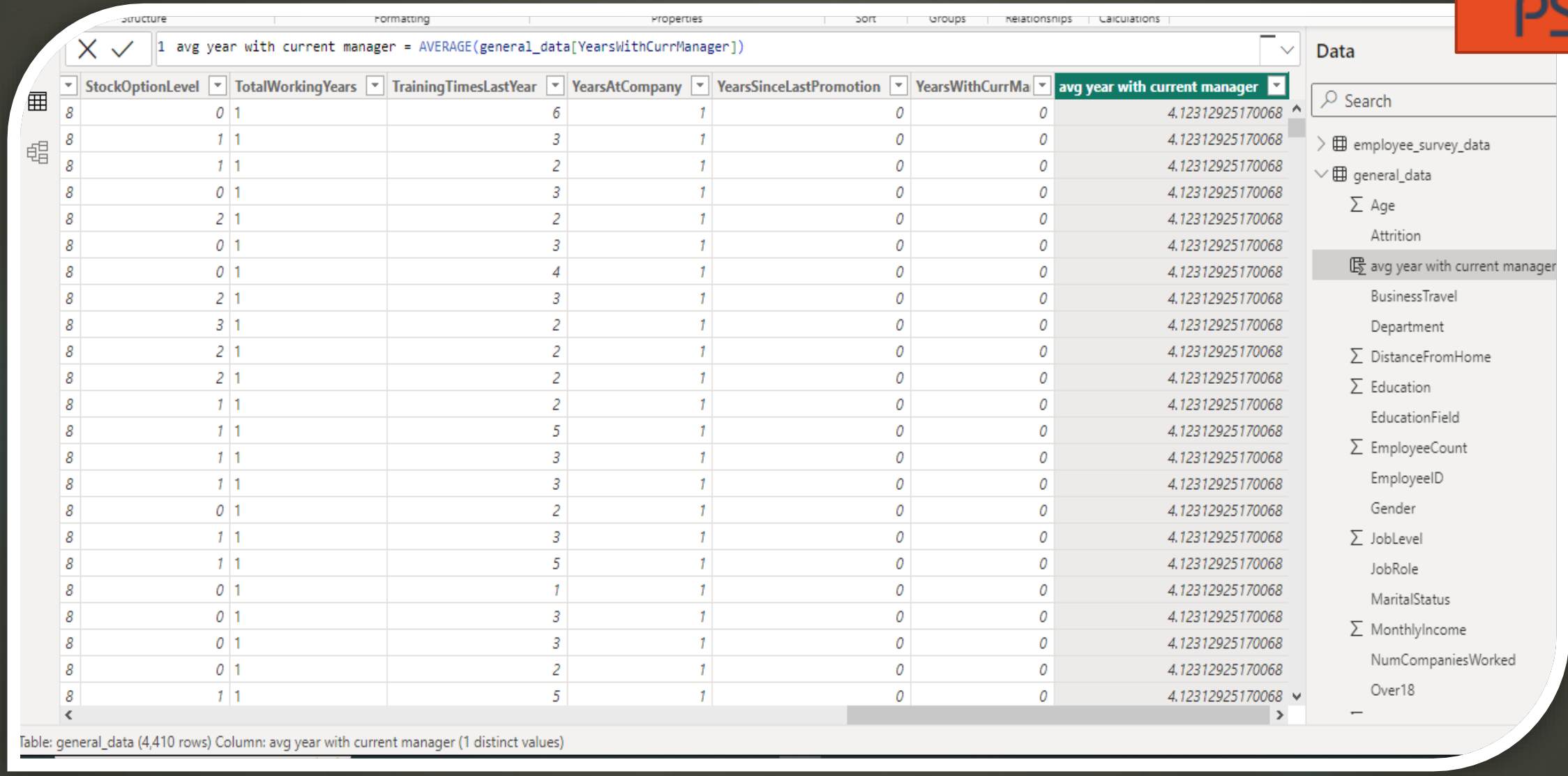
Age	Attrition	BusinessTravel	Department	DistanceFromHome	Educat
51	No	Travel_Rarely		6	
31	Yes	Travel_Frequent		10	
32	No	Travel_Frequent		17	
38	No	Non-Travel		2	
32	No	Travel_Rarely		10	
46	No	Travel_Rarely		8	
28	Yes	Travel_Rarely		11	
29	No	Travel_Rarely		18	
31	No	Travel_Rarely		1	
25	No	Non-Travel		7	
45	No	Travel_Rarely		17	
36	No	Travel_Rarely		28	
55	No	Travel_Rarely		14	
47	Yes	Non-Travel		1	
28	No	Travel_Rarely		1	
37	No	Travel_Rarely		1	
21	No	Travel_Rarely		3	
37	No	Non-Travel		1	
35	No	Travel_Rarely		7	
38	No	Travel_Rarely		8	
26	No	Travel_Frequently	Research & Development	1	
50	No	Travel_Rarely	Sales	8	
53	No	Travel_Rarely	Research & Development	11	

There is no missing and inconsistent data in Department column.

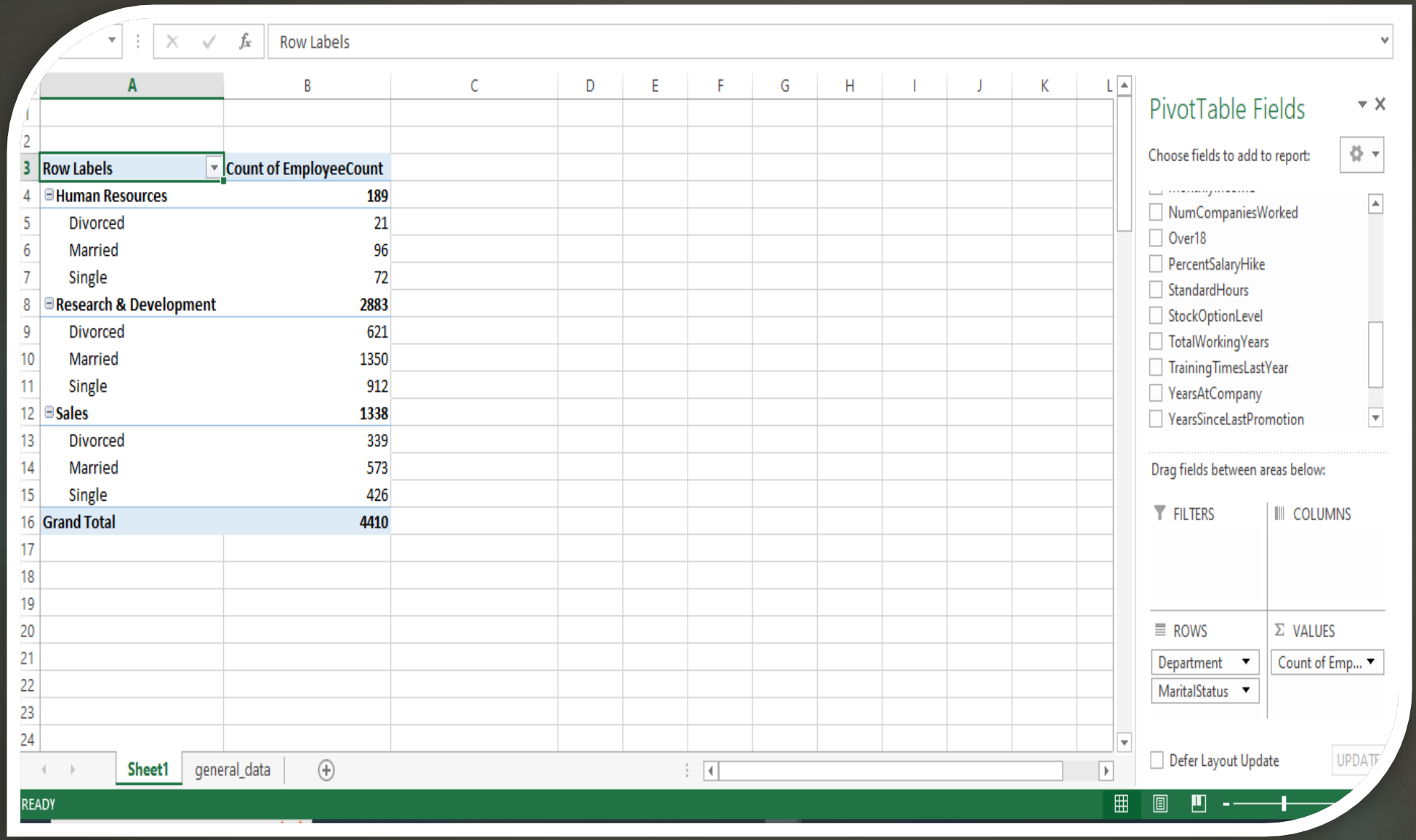
6. In Power BI, establish a relationship between the "EmployeeID" in the employee data and the "EmployeeID" in the time tracking data.-



7. Using DAX, create a calculated column that calculates the average years an employee has spent with their current manager.



8. Using Excel, create a pivot table that displays the count of employees in each Marital Status category, segmented by Department.



Row Labels	Count of EmployeeCount
Human Resources	189
Divorced	21
Married	96
Single	72
Research & Development	2883
Divorced	621
Married	1350
Single	912
Sales	1338
Divorced	339
Married	573
Single	426
Grand Total	4410

PivotTable Fields

Choose fields to add to report:

NumCompaniesWorked
Over18
PercentSalaryHike
StandardHours
StockOptionLevel
TotalWorkingYears
TrainingTimesLastYear
YearsAtCompany
YearsSinceLastPromotion

Drag fields between areas below:

FILTERS | COLUMNS

ROWS: Department, MaritalStatus

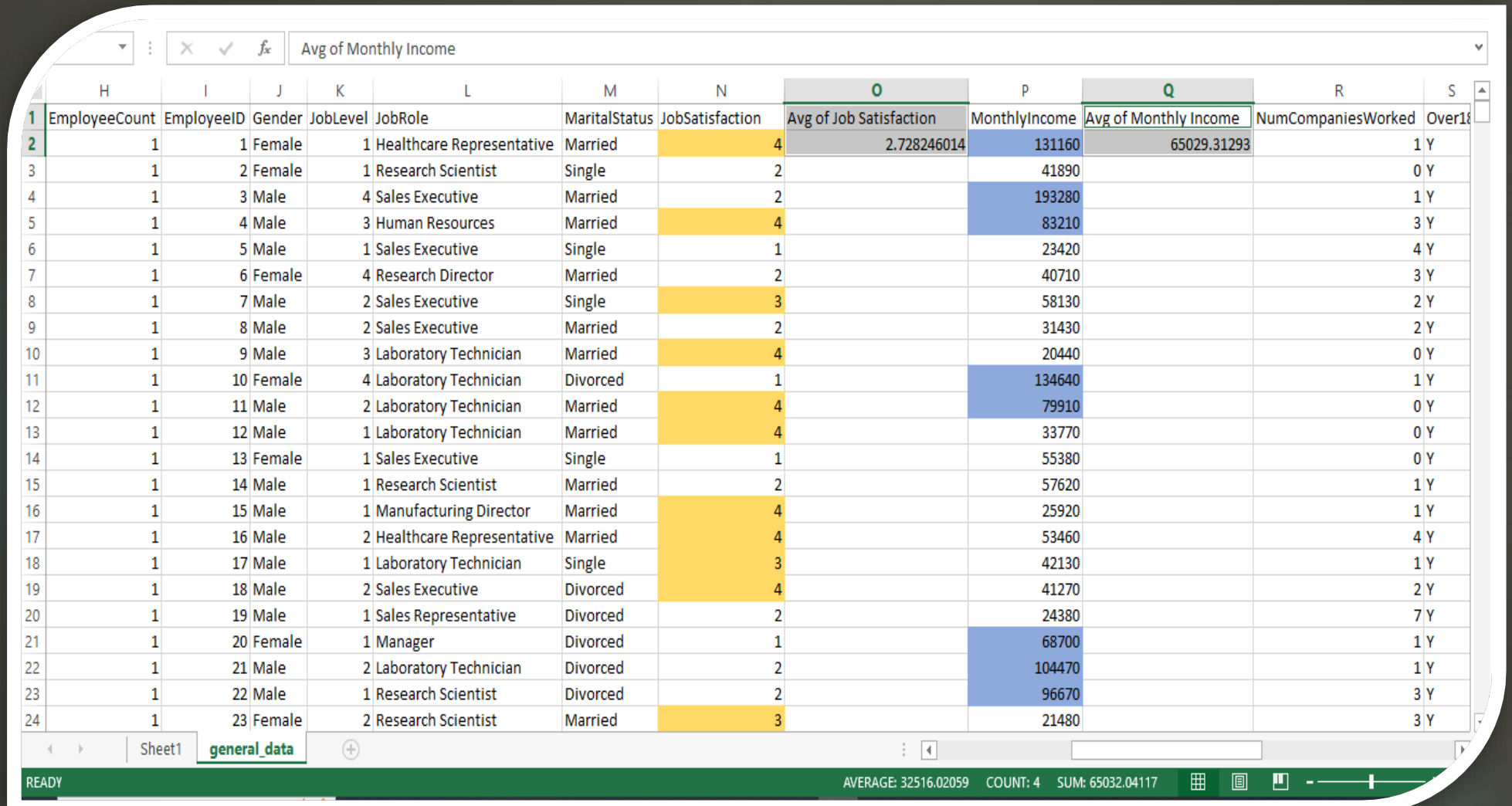
VALUES: Count of Emp...

Defer Layout Update

Sheet1 general_data

READY

9. Apply conditional formatting to highlight employees with both above-average Monthly Income and above-average Job Satisfaction.



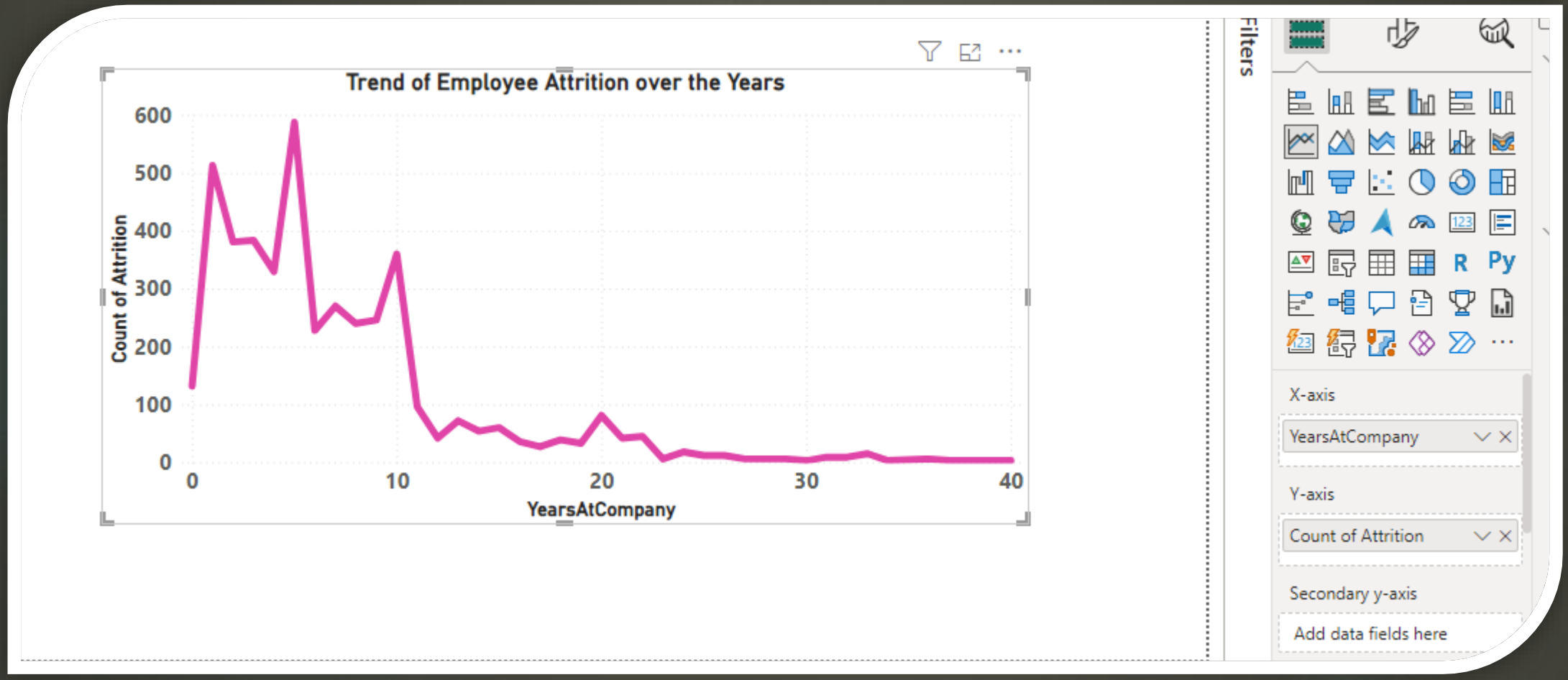
Avg of Monthly Income

	H	I	J	K	L	M	N	O	P	Q	R	S
1	EmployeeCount	EmployeeID	Gender	JobLevel	JobRole	MaritalStatus	JobSatisfaction	Avg of Job Satisfaction	MonthlyIncome	Avg of Monthly Income	NumCompaniesWorked	Over18
2	1	1	Female	1	Healthcare Representative	Married	4	2.728246014	131160	65029.31293	1	Y
3	1	2	Female	1	Research Scientist	Single	2		41890		0	Y
4	1	3	Male	4	Sales Executive	Married	2		193280		1	Y
5	1	4	Male	3	Human Resources	Married	4		83210		3	Y
6	1	5	Male	1	Sales Executive	Single	1		23420		4	Y
7	1	6	Female	4	Research Director	Married	2		40710		3	Y
8	1	7	Male	2	Sales Executive	Single	3		58130		2	Y
9	1	8	Male	2	Sales Executive	Married	2		31430		2	Y
10	1	9	Male	3	Laboratory Technician	Married	4		20440		0	Y
11	1	10	Female	4	Laboratory Technician	Divorced	1		134640		1	Y
12	1	11	Male	2	Laboratory Technician	Married	4		79910		0	Y
13	1	12	Male	1	Laboratory Technician	Married	4		33770		0	Y
14	1	13	Female	1	Sales Executive	Single	1		55380		0	Y
15	1	14	Male	1	Research Scientist	Married	2		57620		1	Y
16	1	15	Male	1	Manufacturing Director	Married	4		25920		1	Y
17	1	16	Male	2	Healthcare Representative	Married	4		53460		4	Y
18	1	17	Male	1	Laboratory Technician	Single	3		42130		1	Y
19	1	18	Male	2	Sales Executive	Divorced	4		41270		2	Y
20	1	19	Male	1	Sales Representative	Divorced	2		24380		7	Y
21	1	20	Female	1	Manager	Divorced	1		68700		1	Y
22	1	21	Male	2	Laboratory Technician	Divorced	2		104470		1	Y
23	1	22	Male	1	Research Scientist	Divorced	2		96670		3	Y
24	1	23	Female	2	Research Scientist	Married	3		21480		3	Y

Sheet1 general_data

READY AVERAGE: 32516.02059 COUNT: 4 SUM: 65032.04117

10. In Power BI, create a line chart that visualizes the trend of Employee Attrition over the years.



11. Describe how you would create a star schema for this dataset, explaining the benefits of doing so.

Steps to Create the Star Schema:

1. **Identify Dimensions and Facts :**
Identify the dimensions (Employee, Manager, Department) and the fact (EmployeePerformance)

2. **Create Primary and Foreign Keys :**
Design primary keys for each dimension table (e.g., EmployeeID, ManagerID, DepartmentID). In the fact table (EmployeePerformance), include foreign keys that reference the primary keys in the dimension tables.

3. **Define Attributes in Dimension Tables :**
Populate dimension tables with attributes relevant to each dimension (e.g., EmployeeName, ManagerName, DepartmentName).

4. **Load Data :**
Load your dataset into the appropriate tables in your database, ensuring data integrity and maintaining relationships between tables.

5. **Create Indexes :**
Implement indexes on primary and foreign keys to optimize query performance.

6. **Test the Schema :**
Run queries to ensure that the schema allows for efficient retrieval and analysis of data. Test joins between the fact and dimension tables.

7. **Document the Schema :**
Document the relationships, primary keys, and foreign keys in your star schema for future reference.

SH

Benefits:

- Star schema is straightforward and easy to understand. It simplifies the structure of the database by separating dimensions and facts, making it easier for users to navigate and query.
- Star schema typically leads to better query performance. Since dimension tables contain descriptive information and are smaller in size, queries can be executed more efficiently.
- Star schema is scalable and flexible. New dimensions or facts can be added without affecting existing structures, facilitating the incorporation of new data and business requirements.
- Maintenance tasks, such as updating or adding data, are simplified. Changes to dimension data (e.g., updating a department name) can be performed without affecting the fact table.
- Query optimization is more feasible in a star schema. Indexing strategies can be applied more effectively to dimension tables, enhancing the speed of retrieval for specific queries.

12. Using DAX, calculate the rolling 3-month average of Monthly Income for each employee.

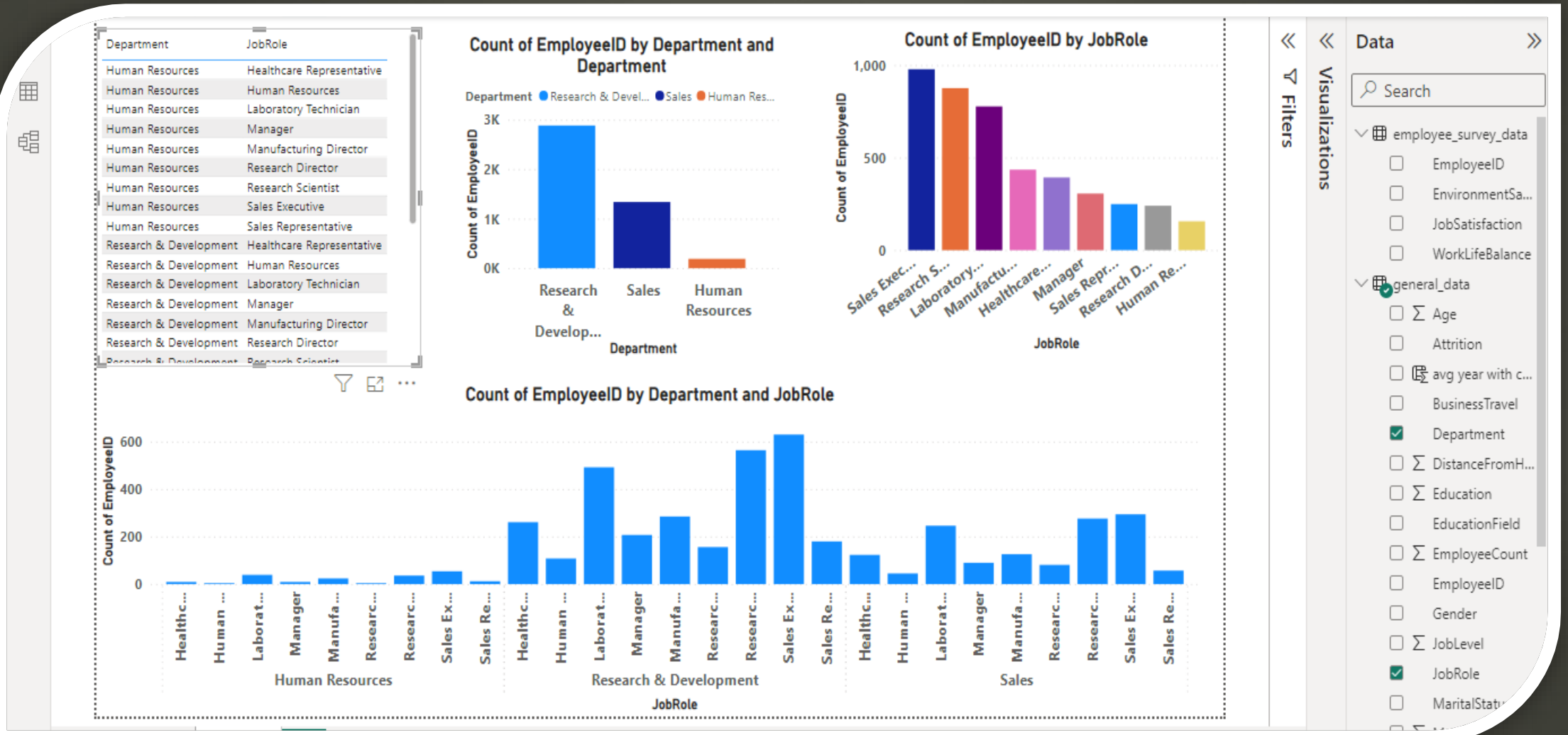
To calculate the rolling 3-month a average of Monthly Income for each employee use following formula,

```
Rolling3MonthsAvgIncome =

CALCULATE( AVERAGE('YourTableName'[MonthlyIncome]),

DATESINPERIOD( 'YourTableName'[Date], LASTDATE('YourTableName'[Date]), -3, MONTH ) )
```

13. Create a hierarchy in Power BI that allows users to drill down from Department to Job Role to further narrow their analysis.



14. How can you set up parameterized queries in Power BI to allow users to filter data based 2 of 2 on the Distance from Home column?

STEPS1: TO CREATE PARAMETER

1.Go to the "Home" ribbon in Power BI Desktop.
2.Click on "Manage Parameters" in the "External Data" group.
3.Click on "New" to create a new parameter and name the parameter (name as Distance)
4.Choose the data type, e.g., Whole Number or Decimal and set a default value.

STEP2 : TO EDIT QUERY

1. select "Transform Data“ in “Home ” ribbon.
2.Locate the table containing the "Distance from Home" column.
3.Click on the drop-down arrow next to the column header and choose "Number Filters"
4.Choose a relevant filter option (e.g., "is equal to," "is greater than," etc.).
5.Instead of providing a static value, reference the parameter.(if the parameter is named "Distance,"
Table.Column = Distance) and click Close and Apply. Close the Power Query Editor and apply the changes.

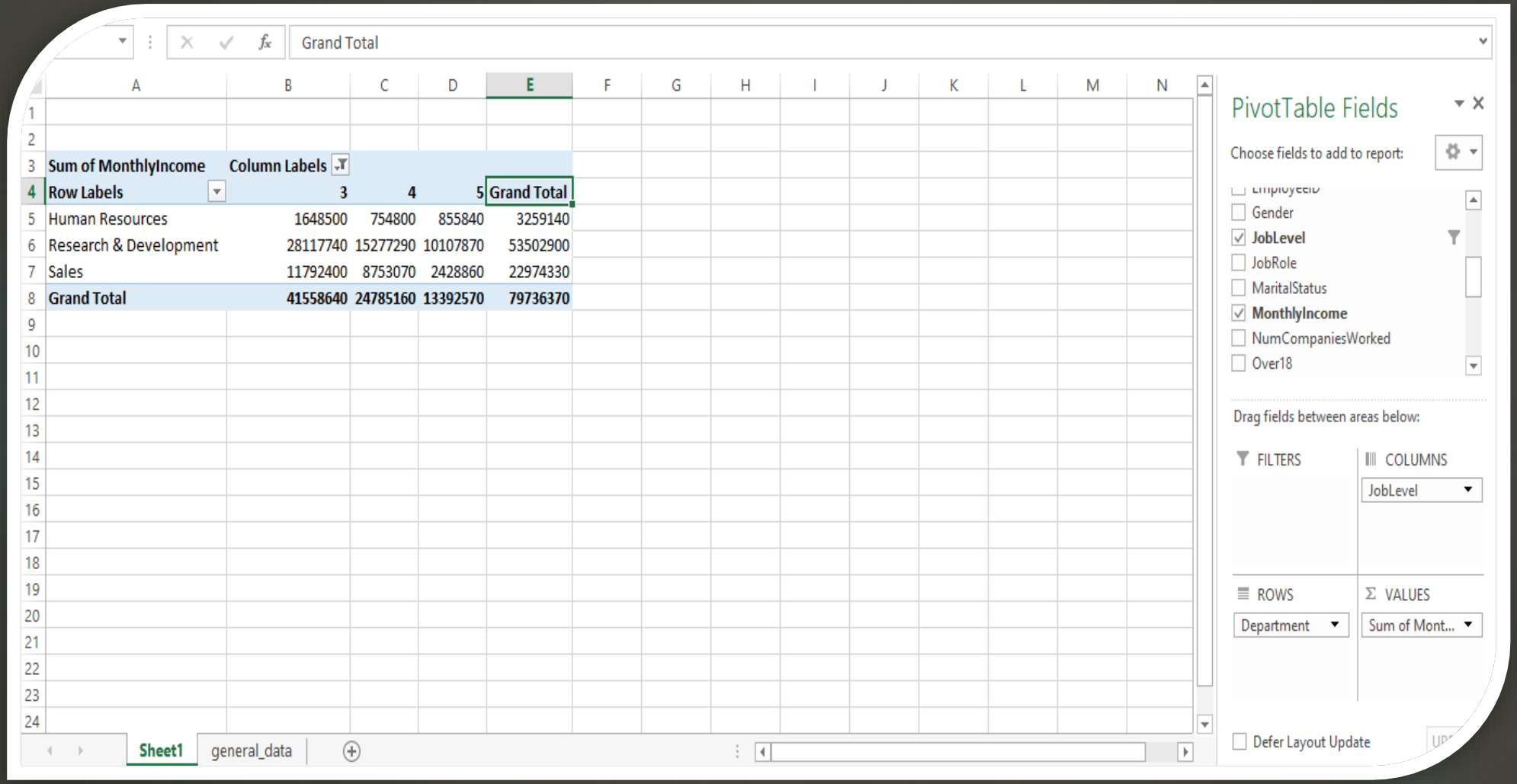
STEP3 : TO CREATE VISUALIZATION

1.Create a slicer visualization based on the "Distance" parameter.
2.We can now adjust the slicer to filter data dynamically based on the chosen distance.

STEP4 : TESTING

1. Test parameterized query in Power BI Desktop to ensure it's working and interact with the parameterized query by adjusting the slicer in the published report.

15. In Excel, calculate the total Monthly Income for each Department, considering only the employees with a Job Level greater than or equal to 3.



16. Explain how to perform a What-If analysis in Excel to understand the impact of a 10% increase in Percent Salary Hike on Monthly Income.

1. Create a table for original and changed values. (as "Original Percent Salary Hike," "Changed Percent Salary Hike," "Original Monthly Income," and "Changed Monthly Income.")

2. In the input table, input the original values for Percent Salary Hike and Monthly Income.

3. In the "Changed Percent Salary Hike" column, input a formula to increase the original Percent Salary Hike by 10%. (=B2 * 1.1 B2 is the original Percent Salary Hike)

4. In the "Changed Monthly Income" column, input a formula to calculate the new Monthly Income based on the changed Percent Salary Hike. (=C2 * (1 + D2) C2 is the original Monthly Income, and D2 is the changed Percent Salary Hike.)

5. Create charts or tables that display the original and changed values side by side for easy comparison.

6. Set up a table with different values for Percent Salary Hike, and use Data Table to automatically calculate Monthly Income for each scenario.

7. Review the results to understand how a 10% increase in Percent Salary Hike affects Monthly Income.

(What-If analysis allows you to explore different scenarios and understand the potential impact of changes in input variables on your outcomes.)

17. Verify if the data adheres to a predefined schema. What actions would you take if you find inconsistencies?

Yes, the data adheres to a redefined schema. Actions to take when Inconsistent are found:

Schema Comparison: Check the dataset's columns, data types, and relationships against the predefined schema. For instance, if the schema dictates a column should contain numerical values and it contains text, that's an inconsistency.

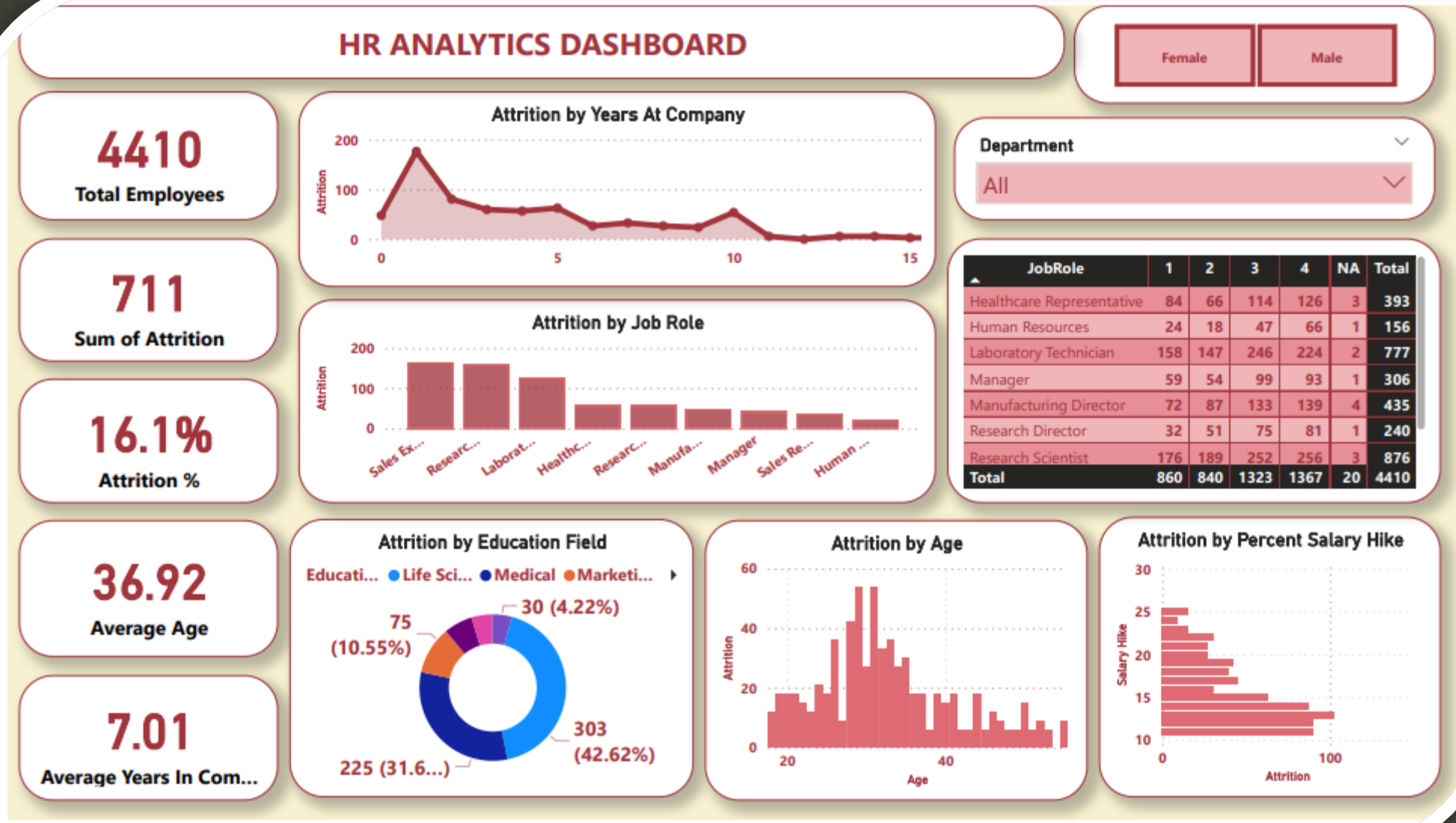
Inconsistency Resolution: Data Cleaning: Correct minor inconsistencies by cleaning or transforming the data. For example, converting text into the expected numerical format.

Imputation: Fill missing values using appropriate methods like mean, median, or using domain-specific knowledge.

Normalization: Standardize data formats or values according to the schema. This might involve ensuring date formats are uniform or converting units of measurement.

Documentation and Communication: Document the identified inconsistencies and the actions taken. Communicate these findings to relevant stakeholders or data owners for validation or further instructions.

REPORT


HR ANALYTICS DASHBOARD
Female
Male
4410
Total Employees
711
Sum of Attrition
16.1%
Attrition %
36.92
Average Age
7.01
Average Years In Com...
Attrition by Years At Company
Attrition
200
100
0
0
5
10
15
Attrition by Job Role
Attrition
200
100
0
Sales Ex...
Researc...
Laborat...
Healthc...
Researc...
Manufa...
Manager
Sales Re...
Human ...
Department
All
JobRole 1 2 3 4 NA Total
Healthcare Representative 84 66 114 126 3 393
Human Resources 24 18 47 66 1 156
Laboratory Technician 158 147 246 224 2 777
Manager 59 54 99 93 1 306
Manufacturing Director 72 87 133 139 4 435
Research Director 32 51 75 81 1 240
Research Scientist 176 189 252 256 3 876
Total 860 840 1323 1367 20 4410
Attrition by Education Field
Educati... Life Sci... Medical Marketi...
30 (4.22%)
75
(10.55%)
303
(42.62%)
225 (31.6...)
Attrition by Age
Attrition
60
40
20
0
20
40
Age
Attrition by Percent Salary Hike
Salary Hike
30
25
20
15
10
0
100
Attrition